BOX PATENT APPLICATION
Assistant Commissioner for
Patents
Washington, D.C. 20231

CERTIFICATE OF MAILING BY EXPRESS MAIL

"EXPRESS MAIL" Mailing Label No. EL 798306573US

Date of Deposit: July 19, 2001

I hereby certify that this paper or fee is being deposited with the U.S. Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231

Type or Print Name CAROL YOUNG

Carol Young

Signature

DYNAMIC QoS MANAGEMENT IN DIFFERENTIATED SERVICES USING BANDWIDTH BROKERS, RSVP AGGREGATION AND LOAD CONTROL PROTOCOLS

This application claims the priority under 35 U.S.C. 119(e)(1) of copending U.S. Provisional Application No. 60/221,773, filed on July 31, 2000, which is incorporated herein by reference.

Field of the Invention

This invention relates generally to providing Quality of Service (QoS) in Internet Protocol networks, and more particularly it relates to providing dynamically and on demand end to end QoS in Internet Protocol networks using RSVP aggregation and load control.

The text herein makes reference to several protocols and standard-like publications which include, but are not limited to, the following RFC numbers: 1633, 1905, 2748 2205, 2210, 2475, 2597, 2598, 2638, 2748 and 2998. A complete list of references relied upon herein is attached at the end of the text.

BACKGROUND OF THE INVENTION

The diversity of the current Internet applications from the most simple ones like e-mailing and web browsing going to high demanding real time applications like IP telephony and multimedia conferencing, has raised the expectations that both users and software developers of these applications have from the Internet. On the other hand, in such a highly competitive environment as the Internet Service Providers' (ISPs) world, satisfying customer needs, regardless of whether they are other ISPs or end users is key to their survival. Therefore the ISP's zeal to provide value-added services to their customers is growing immensely. These demands have led to evolution of QoS on Internet as a necessity. Enabling QoS on the best effort Internet model introduces complexity in several aspects, starting from not only applications, different networking layers and network architectures but also in network management and business models. All these aspects have been major research

topics over the past few years. Finding an efficient solution for end-to-end QoS over Internet i.e. the IP networks that will satisfy both ISPs and their customers is a tough venture. The efforts to enable end-to-end QoS over IP networks have led to the development of two different architectures, the Integrated Services architecture and more recently, the Differentiated Services architecture.

INTEGRATED SERVICES ARCHITECTURE

Integrated Services (Intserv) architecture [RFC 1633] uses an explicit mechanism to signal per-flow QoS requirements to network elements (hosts, routers). Network elements, depending on the available resources, implement one of the defined Intserv services (Guaranteed or Control Load service) based on which QoS will be delivered in the data transmission path. Another protocol known as the RSVP signaling protocol [RFC2205], [RFC2210] was designed as a dynamic mechanism for explicit reservation of resources in Intserv, although Intserv can use other mechanisms as well. It is initiated by an application at the beginning of a communication session. But, even though Intserv is deigned to provide end-to-end QoS it is currently not widely deployed, as fairly well known now, due to maintenance and control of per-flow states and classification; reserving resources per-flow introduces severe scalability problems at the

core networks, where the number of processed flows is in the millions range. Consequently, the use of the Integrated Services architecture is limited to small access networks where the number of flows using reservations is modest.

RSVP

The Resource Reservation Protocol (RSVP) (which is explained in references [RFC2210], [DuYa99]) is a signaling protocol that can be used by an application to inform its QoS requirements to an Internet infrastructure. RSVP is initiated by an application at the beginning of a communication session. A communication session is identified by a combination of the IP destination address, transport layer protocol type and the destination port number. 'RSVP' as used herein is meant to include any and all modifications of what is known as RSVP.

A simplified RSVP/Intserv framework is illustrated in Figure 1. As shown, every RSVP aware router in the Intserv will be able to perform RSVP signaling, admission control, policing and scheduling.

The resources reserved by the RSVP for a certain communication session will be used for all packets belonging to that particular session. Therefore, all RSVP packets will include details of the session they belong to. The main RSVP messages are the PATH and RESV messages. The PATH message is

sent by a source that initiates the communication session and it explicitly binds the data path of a flow. Furthermore, it describes the capabilities of the source. The RESV message is issued by the receiver of the communication session and it follows exactly the path that the RSVP PATH message has followed hop by hop back to the communication session source. The RESV message on its way back to the source, may install QoS states at each hop. These states are associated with the specific QoS resource requirements of the destination. The RSVP reservation states are temporary states, i.e., soft states, that have to be updated regularly. This means that PATH and RESV messages will have to be periodically retransmitted. If these states are not refreshed then they will be removed.

The RSVP protocol uses additional messages that are used to either provide information about the QoS state or to explicitly delete the QoS states along the communication session path. This invention envisages the use and application of RSVP protocol.

The RSVP message functions and their meaning are given in Table 1.

RSVP Messages	
RSVP Message Name	**RSVP Message Function**
PATH	The PATH message is sent by a source 101 (figure 1) that initiates the communication session to destination 102 and it explicitly binds the data path of a flow. Furthermore, it describes the capabilities of the source.
RESV	The RESV message is issued by the receiver of the communication session and it follows exactly the path that the RSVP PATH message has followed hop by hop back to the communication session source. The RESV message on its way back to the source may install QoS states at each hop. These states are associated with the specific QoS resource requirements of the destination. The RSVP reservation states are temporary states, i.e., soft states, that have to be updated regularly.
PATH Error	Is used to report errors that are occurring during the installation of a path from the source to the destination 102, in figure 1, of a communication session.
RESV Error	Is used to report errors that are occurring during the installation of the reservation states along the communication session path.
RESV Confirm	It provides a positive indication to the initiator of the communication session informing that all nodes along the communication session path accepted the reservation request. The RSVP Confirmation messages are typically sent by the source of the communication session directly to the destination of this communication session. Intermediate nodes do not process RSVP confirmation messages.
PATH Tear	Is sent by the source of the communication session and it explicitly deletes the stored QoS state information on all nodes included in a communication session path.
RESV Tear	Is sent by the destination of the communication session and it explicitly deletes the stored QoS state information on all nodes included in a communication session path.

Table 1: The RSVP messages

DIFFERENTIATED SERVICES ARCHITECTURE

Differentiated Services (Diffserv) architecture [RFC2475], [RFC2638], was introduced as a result of the efforts to avoid the scalability and complexity problems of Intserv.

In Diffserv, per-flow state is pushed to the edges, and the traffic through Diffserv routers is treated on an aggregate basis. Service differentiation is achieved by means of Differentiated Service (DS) field in the IP header and the Per-Hop Behaviour (PHB) as main building blocks. At each node, packets are handled according to the PHB invoked by the DS byte in the packet header. The PHB defines the externally observable behaviour at the node. Two PHBs have been defined, the assured forwarding (AF-) PHB [RFC2597] and the expedited forwarding (EF-) PHB [RFC2598]. The Diffserv domain will provide to its customer, which is a host or another domain, the required service by complying fully with the agreed Service Level Agreement (SLA). SLA is a bilateral agreement between the boundary domains negotiated either statically or dynamically. The transit service to be provided with accompanying parameters like transmit capacity, burst size and peak rate, is specified in the technical part of the SLA, i.e. the Service Level Specification (SLS), shown at 203 in figure 2. The Diffserv architecture is certainly promising, but there are several open

issues related to intra-domain resource allocation mechanisms and inter-domain communication in case of dynamic resource provisioning that need to be defined and researched. The simplified Diffserv framework is generally illustrated in Figure 2.

The Diffserv architecture may use different types of protocols to dynamically reserve resources into the Diffserv domain. The main ones are the RSVP, RSVP aggregation and Load Control protocols.

RSVP Aggregation

The RSVP aggregation concept is used to reduce the Intserv scalability problems by extending the RSVP protocol with facilities for aggregation of individual reserved sessions into a common class and across transit domains. It describes mechanisms for dynamic creation of the aggregate reservation, classification of the traffic for which the aggregate reservation applies, determination of the bandwidth needed to achieve the requirement and recovery of the bandwidth when the sub-reservations are no longer required. An RSVP aggregated session is identified by the IP destination address, the protocol ID and Differentiated Services Code Points (DSCP) information. Furthermore, for classification and scheduling of traffic supported by aggregate reservations, Diffserv mechanisms

are used. Diffserv DSCPs are used to identify traffic covered by aggregate reservations and, one or more Diffserv PHB's are used to offer the required forwarding treatment to this traffic.

The first router in the transit domain that handles the aggregated reservations is called Aggregator, while the last router in the transit domain that handles the reservations is called Deaggregator. An RSVP aggregation region comprises routers that are capable of managing the RSVP aggregated states.

The RSVP aggregation concept can be used either when RSVP is applied end to end or edge to edge. In the latter case the Aggregator can use a policy that can be based on local configurations and local QoS management architectures, to set the DSCP packets that are passing into the aggregated region. For example, the Aggregator may be a PSTN (Public Switched Telephone Network) gateway that aggregates a set of incoming calls and makes an aggregate reservation across one or more Diffserv domains up to the Deaggregator that can be e.g., another PSTN gateway. In this situation, call signaling is used to establish the E2E reservations.

From the above information it can be deduced that based on a certain policy the Aggregator and Deaggregator will decide when the RSVP Aggregated states will be refreshed or updated and therefore this triggering time is not completely defined by the E2E RSVP messages.

Within an aggregation region three possible scenarios can be distinguished.

SIGNALING FLOW FOR THE FIRST E2E FLOW

The flow diagram depicted in Figure 3 illustrates a detailed flow of RSVP messages in the case when there is no Aggregated PATH between the Aggregator 301 and the Deaggregator 303. Also shown is the location of an Intermediate Router 302. The number of the Aggregated PATH states that will have to be installed in each router depends on the number of the supported DSCPs. In figure 3 it is considered that two DSCPs are supported, e.g., EF and AF PHB's, and therefore, two RSVP PATH aggregated states have to be created. The symbols (A) and (B) in the flow diagram represent new aggregation values needed for the different supported DCSP's. The E2E RSVP messages transport the value that identifies a particular DSCP (e.g., PHB) type in the DCLASS object.

SIGNALING FLOW FOR SUBSEQUENT E2E FLOW WITHOUT RESERVATION RESIZING

Figure 4 illustrates a detailed flow of RSVP messages in the case that there already exists an Aggregated PATH between the Aggregator 401 and Deaggregator 403, and there is no need for a change in the RSVP aggregated reservation. The E2E RSVP

messages transport the value that identifies a particular DSCP (e.g., PHB) type in the DCLASS object.

SIGNALING FLOW FOR SUBSEQUENT E2E FLOW WITH RESERVATION RESIZING

Figure 5 illustrates a detailed flow of RSVP messages in the case when there already exists an Aggregated PATH between the Aggregator 501 and Deaggregator 503 and there is a need for a change in the RSVP aggregated reservation © - represents new values, e.g. more bandwidth). The E2E RSVP messages transport the value that identifies a particular DSCP (e.g., PHB) type in the DCLASS object. Also shown is an intermediate router 502.

SIGNALING FLOW FOR E2E FLOW RELEASE

The flow diagram depicted in Figure 6 illustrates an E2E RSVP release procedure. Illustrated in Figure 6 are an Aggregator 601, deaggregator 603 and an intermediate router 602.

LOAD CONTROL

Load control is a scheme for resource allocation within the Diffserv networks, without requiring out-band signaling or any per-flow processing in core routers. "Load control scheme" as used herein is to be understood as any available load control protocol or equivalent schemes. A load control scheme is

generally designed so as to perform admission control of incoming request and to drop the admitted flows in case of failure events, e.g. link failure. Load Control related information can be stored in the Diffserv packet headers by using either new Differentiated Services Code Points (DSCP) or using the two least significant bits of either the IPv4 TOS (Type of Service) header octet or the Ipv6 Traffic Class header octet. The Load control scheme has two modes of operation, i.e., simple marking, which uses resource probing, and unit-based reservation.

Figure 7 shows a view of the basic operation of the Load control unit based reservation scheme. The resource reservation, during one refreshment period (i.e. period (i)), can be achieved by sending probe packet (PP) or refreshed packet (RP) messages. If a router changes the PP or RP messages into a marked packet (MP), it means that the resource reservation procedure for that unit of resource could not be accomplished. If these messages are not changed into MP messages then it means that the resource reservation procedure has been able to reserve the resources for that unit of resource during period (i+1). The resource release procedure during a period can be achieved when the resource reservation mechanism does not send any PP or RP messages, but only ordinary packet (OP) messages.

Illustrated in Figure 7 are Terminal 1 (701), Terminal 2 (704), Border Router 1 (702) and Border Router 2 (703).

In Figure 7 the Load Control operation is accomplished by providing the possibility to the Ingress/Egress Border Routers 702, 703 to use the Normal IP packets 705 that are sent by the end Terminals 701,704 and to mark them as probe packets (PP), refreshed packets (RP), ordinary packets (OP) or marked packets (MP) packets.

However, there are situations that the Ingress/Egress Border Routers will not receive any Normal IP packets. For these situations it is noted that the Ingress/Egress Border Routers should be able to generate dummy IP packets, i.e., IP packets without a payload. These dummy IP packets will be then used as PP, RP, OP or MP packets.

DRAWBACKS IN PRIOR ART ARRANGEMENTS

The Internet environment is a highly competitive environment where different IP network operators need to satisfy the customer demands for quality in the supported applications. Finding an efficient solution for end to end QoS over Internet that will satisfy both IP network operators and the demands/needs of their customers is a real challenge. A promising solution to this challenge is the expedient use of the dynamic provisioning of QoS in the IP networks. Dynamic QoS

provisioning may be defined as a procedure wherein the QoS provided by a network can be initiated or modified instantaneously either on the demand of an application or based on a predefined procedure described in a service profile.

The IP network operators using dynamic QoS should be satisfied since they will obtain the ability of controlling the utilization of their network instantaneously. Notwithstanding, there may still be some concerns and unsolved issues which are related to the end to end QoS dynamic provisioning in IP networks. The most important issues include:

1. **Issue_1:** The end to end QoS demands of end users should be satisfied.
2. **Issue_2:** The QoS management architectures used in the core network must be scalable.
3. **Issue_3:** The different QoS management architectures that are combined and used in the end to end communication must easily interoperate.

KNOWN SOLUTIONS

Currently three main solutions can be used to solve one or more of the three issues listed supra.:

4. **Solution_1:** The current Intserv architecture described earlier can provide efficient solutions to the **Issue_1** and **Issue_3.**

5. **Solution_2:** The current Diffserv architecture described earlier can provide efficient solutions to **Issue_3**. **Issue_1** and **Issue_2** can be partially solved.

6. **Solution_3:** The QoS management architecture, called Intserv over Diffserv, can provide efficient solutions to **Issue_1**. **Issue_2** and **Issue_3** can be partially solved. This QoS management architecture combines the Intserv and Diffserv QoS architectures and uses them as complementary technologies in the access and the core networks respectively. The main functionality for the Intserv over Diffserv operation will be performed at the edge devices either at Intserv or Diffserv, i.e., Edge Routers (ER1, ER2) and Border Routers (BR1, BR2), respectively, depending on the specific configurations of the framework. These devices will have the burden of handling both RSVP/Intserv messages and Diffserv packets.

Figure 8 generally illustrates a reference network for the RSVP/Intserv over Diffserv proposed framework. Figure 8 shows a sender 801, a receiver 802, non-Diffserv regions 803, 804, respectively including Edge Routers ER1 (805) and ER2 (806), and a Diffserv region 807 having Border Routers BR1 and BR2 (808). The dynamic QoS management can be accomplished by RSVP aware

Border Routers and Core Routers in each Diffserv domain by using per flow, tunneled, aggregated RSVP.

- **Solution_4:** An enhancement of the Intserv over Diffserv framework is proposed wherein the RSVP aggregation protocol described supra is used to reserve aggregated resources on some of the Border Routers and Core routers of each Diffserv domain. This framework is depicted in Figure 9. Figure 9 shows a sender 901, receiver 902, Intserve regions 903,904 and Diffserv regions 905. A Border Router, i.e., BR, can operate as an Aggregator and Deaggregator as described hereinabove. The dynamic QoS management architecture is accomplished by RSVP aggregation aware Border Routers and Core Routers in each Diffserv domain.

However, there are problems and deficiencies with the afore-mentioned solutions.

Due to the fact that the resource reservation states stored in all the RSVP aware Border and Core routers are representing aggregated RSVP sessions (i.e., trunks of RSVP sessions), the scalability problems on the routers will be drastically minimized. However, it is believed that in a full meshed Diffserv network, as shown for example in **Figure 10**, the number of the RSVP aggregated sessions grows as: number_aggregates = $n^2 - n$, where n represents the number of Border Routers 1001 that simultaneously send and receive information to/from all the

other Border Routers of the Diffserv domain 1002. For example, in Figure 10 where the number of Border Routers is n=3, the maximum number of simultaneous RSVP aggregated sessions is *number*_aggregates=6. This means that the number of the aggregated states that each Core Router will have to simultaneously maintain is increasing with the number of the Border Routers (=n) that simultaneously send and receive information to/from all the other Border Routers 1001 of the Diffserv domain using the equation:

number_aggregates=n^2 - n. When the number of the Border Routers 1001 is high, e.g., 200 then: *number_aggregates* = 39800. This may cause scalability problems on the Core Routers of the Diffserv domain. Therefore, it can be concluded that **Solution_4** will solve **Issue_2** only for small Diffserv domains. When the Diffserv domain is large, i.e., including many Border Routers, then **Solution_4** will not be able to solve **Issue_2**.

- **Problems with Issue_3**: In **Solution_4** the dynamic QoS management architecture is accomplished by the RSVP aggregation aware Border Routers and Core Routers in each Diffserv domain. Due to the fact that not all the Border and Core Routers will be RSVP aggregation aware, it will be difficult to interoperate, maintain and manage the end to end QoS provisioning.

Therefore, **Solution_4** will not be able to efficiently solve **Issue_3**.

It is seen in light of the above discussions that QoS solutions offered by hitherto known arrangements have setbacks and disadvantages whereby there still exists a need for more efficient provisioning of dynamic on demand end to end QoS in IP networks.

BRIEF SUMMARY OF THE INVENTION

The present invention overcomes the disadvantages posed by prior art solutions to providing dynamic QoS. The invention ensures dynamic end to end QoS in IP networks using a judicious combination of RSVP aggregation, Load Control and Bandwidth brokers used selectively, which can operate on a predetermined protocol. "Bandwidth" brokers in this invention are connected and made to work differently from bandwidth brokers known in prior art.

This invention enhances and extends the Intserv over Diffserv framework that was used in **Solution_4** and described supra under the subtitle KNOWN SOLUTIONS, such that all the three issues, **Issue_1, Issue_2** and **Issue_3** described earlier under the subtitle "Discussion of Drawbacks" are efficiently addressed and solved.

The invention offers a new framework that is an enhancement and extension of the Intserv over Diffserv framework used in **Solution_4** described earlier. This new framework is able to efficiently solve **Issue_1, Issue_2** and **Issue_3** described earlier under the subtitle "Discussion of Drawbacks."

The following are preferred requirements which the present invention is intended to satisfy:

- **Requirement_1:** The QoS dynamic provisioning in each Diffserv domain must be arranged by a Bandwidth Broker (BB) with a predetermined functionality.
- **Requirement_2:** The IP core network is based on either the Diffserv network architecture or a mix of Diffserv and overprovisioned IP core networks. The second option will be valid especially if the provider of IP networks guarantees certain QoS bounds.
- **Requirement_3:** First of all a new functionality for the Bandwidth Broker (BB) entity used in a Diffserv network is introduced. The BB is currently specified as a centralized oracle that has sufficient knowledge of resource availability and network topology to make admission control decisions. It has been discovered that if the BB is used to obtain the resource availability in the Diffserv domain by directly querying all the Border and Core Routers in the

Diffserv domain, this will impose severe scalability problems into the BB. Therefore, a modified way is to be specified that will enable the BB to obtain the required resource availability of the resources but will not introduce severe scalability problems into the BB.

The modification is related to the fact that the BB is made to directly communicate, by using e.g., Common Open Policy Service or Simple Network Management (COPS or SNMP) protocols, and manage only the Border Routers and not the Core Routers in the Diffserv Domain. In this way the BB is able to request from all the ingress BR's to either reserve a certain amount of resources or refresh a reservation that has been accomplished during a previous refreshment period.

- **Requirement_4**: Each BB must be able to use the RSVP aggregation protocol described supra. Furthermore, each BB must be able to store and manage the RSVP aggregation states.
- **Requirement_5**: The Border Routers must manage the resource availability and the admission control into the interior of the Diffserv domain, i.e., on the Core Routers. This can be achieved by using the Load Control protocol described earlier.
- **Requirement_6**: The Intserv to Diffserv interoperability must be accomplished using one of the following ways:

- by the Edge Router (shown in Figure 11) and either the Bandwidth Broker Aggregator (BBA) or the Bandwidth Broker Deaggregator (BBD). If any Border Router or Core Router is RSVP aware, then IP tunneling may be used to avoid the situation that the applied RSVP messages, i.e., E2E RSVP or aggregated RSVP, will be processed by the Border or Core routers.
  - by the Edge Router, the first Border Router and either the BBA or the BBD. In this situation the first Border Router in a Diffserv domain that can communicate with either a BBA or a BBD and an Edge router must be RSVP aware, then IP tunneling may be used to avoid the situation that the applied RSVP messages, i.e., E2E RSVP or aggregated RSVP, will be processed by these Border or Core routers.

The invention in its broad form resides in a network subsystem for providing dynamic quality of service (QoS) in an IP network which handles IP packets, the network using Resource Reservation Protocol (RSVP) aggregation and differentiated services architecture (Diffserv) including at least one Diffserv domain, said system comprising a bandwidth broker (BB) which manages dynamic provisioning of QoS in each Diffserv domain,

using a predetermined protocol, and storing RSVP aggregated states in the bandwidth broker.

In another aspect, the invention resides in a network subsystem for providing dynamically and on demand end to end quality of service in an IP network which uses Resource Reservation Protocol (RSVP) aggregation and differentiated services architecture (Diffserv), said Diffserv comprising at least one Diffserv domain including Border Routers (BRs) and Core Routers (CRs), said network subsystem comprising: a Bandwidth Broker (BB) which stores aggregated states and manages dynamic provisioning of QoS in each Diffserv domain using a predetermined protocol.

The invention also resides in a method of providing dynamic quality of service (QoS) in an IP network which handles IP packets and being of the type which uses RSVP (Resource Reservation Protocol) aggregation and differentiated services architecture (Diffserv), said Diffserv comprising a Diffserv domain including Border Routers (BR) and Core Routers (CR), said method comprising the steps of:

managing dynamic provisioning of QoS in each Diffserv domain by using a bandwidth broker (BB) which communicates using a predetermined protocol.

In a modification, the invention resides in a method, in an IP network of the type which handles IP packets and uses

Resource Reservation Protocol (RSVP) aggregation and differentiated services architecture (Diffserv), said Diffserv comprising a Diffserv domain including Border Routers (BR) and Core Routers (CR), said method providing end to end quality of service (QoS) on demand, said method comprising managing dynamic provisioning of QoS in each Diffserv domain by using a bandwidth broker (BB) which communicates using a predetermined protocol.

The predetermined protocol may be for instance what is known as Common Open Policy Service Protocol (COPS) or may be a Simple Network Management Protocol (SNMP).

In a modification of the method of the invention, the step of managing comprises using a BB which obtains resource availability information by communicating only with BRs to the exclusion of CRs.

Yet another modification of the inventive method, includes the step of reserving resources through a BB which queries BRs.

A further modification of the inventive method includes the step of refreshing a reservation of resources, which reservation has been accomplished during a previous refreshment period.

In a variation of the inventive method, the BB is capable of using an RSVP aggregation protocol and is able to store and manage RSVP aggregation states.

A further variation of the inventive method includes the step of using Load Control Protocol, and managing, by use of a

BR, resource availability and admission control into an interior of said Diffserv domain.

In a modification of the inventive network subsystem the Diffserv domain includes Border Routers (BRs) and Core Routers (CRs), and the BB obtains resource availability information by communicating with BRs.

In yet another modification, the BB reserves resources by querying BRs to the exclusion of CRs; alternatively, the BB refreshes an already made reservation of resources which reservation has been accomplished during a previous refreshment period.

In a variation of the inventive network subsystem, the BB is capable of using an RSVP aggregation protocol and is able to store and manage RSVP aggregation states.

In yet another variation, load control protocol is used, and, by the use of a BR, resource availability and admission control into an interior of a Diffserv domain are managed.

BRIEF DESCRIPTION OF THE DRAWINGS

A more detailed understanding of the invention may be had from the following description of preferred embodiments, given by way of example and to be understood in conjunction with the accompanying drawing, wherein:

FIGURE 1 illustrates a simplified RSVP/Intserv framework;

FIGURE 2 illustrates a simplified Diffserv framework;

FIGURE 3 illustrates a detailed flow of E2E RSVP messages;

FIGURE 4 illustrates RSVP aggregation flow for subsequent E2E flow without Reservation resizing;

FIGURE 5 illustrates RSVP aggregation signaling for subsequent E2E flow with reservation resizing;

FIGURE 6 illustrates an E2E RSVP release procedure;

FIGURE 7 illustrates resource reservation and resource release procedures in load control;

FIGURE 8 illustrates the reference network for the RSVP/Intserv over a Diffserv framework;

FIGURE 9 illustrates Intserv over Diffserv framework using RSVP aggregation within each Diffserv domain;

FIGURE 10 shows an example of a full meshed Diffserv domain with three border routers (BR) and one core router (CR);

FIGURE 11 illustrates a proposed inventive subsystem of the network using an Intserv/Diffserv framework and a bandwidth broker;

FIGURE 12 illustrates a proposed inventive subsystem of the network using Intserv/Diffserv operation when RSVP aggregated states are not available in the BBs;

FIGURE 13 illustrates a proposed inventive subsystem of the network using Intserv/Diffserv operation when RSVP aggregated states are available in the BBs, and no resizing is needed;

FIGURE 14 illustrates a modification of the arrangement in Figure 13 where resizing is needed; and

FIGURE 15 shows an example of refreshment of reserved resources.

DETAILED DESCRIPTION OF THE INVENTION

Described hereinafter are an exemplary method and network subsystem which use a combination of bandwidth brokers, RSVP aggregation and load control protocols, to achieve a dynamic end to end QoS.

<u>Operation</u>

The QoS dynamic provisioning mechanism in a Diffserv domain can use a resource reservation protocol that will be able to inter-communicate with the QoS mechanisms applied in neighboring domains (Diffserv or non-Diffserv). Such a protocol can be the RSVP aggregation protocol described earlier but preferably with a modification. The modification can consist, for example, in that the Border Routers (BR) will not anymore store the RSVP aggregated states, but these states will be stored in the Bandwidth Brokers (see **Requirement_4**).

Figure 11 generally illustrates a preferred embodiment of the invention showing sender 1101 and receiver 1102 between which end to end QoS is intended to be achieved. As shown, Edge Router ER1 shown at 1103 interacts with sender 1101, and Edge

Router ER2 shown at 1102 interacts with receiver 1102. Also shown in Figure 11 are Intserv regions 1108, and Diftserv regions 1109 interacting with Bandwidth Broker Aggregator BBA/D, shown at 1105, Bandwidth Broker BB shown at 1106, and Bandwidth Broker Aggregator BBD/A shown at 1107. Aggregated RSVP messages 1110 flow between 1105, 1106, and 1106, 1107.

As shown Figure 11 the network element 1105 (BBA) is functioning as a Bandwidth Broker Aggregator and the nework element 1104 (BBD) is functioning as a Bandwidth Broker Deaggregator.

Furthermore, the E2E RSVP and RSVP aggregation messages in Figure 11 are exchanged between the Intserv networks and the BB's that are managing each Diffserv domain and are used to provide the QoS end to end provisioning. The Intserv to Diffserv interoperability (see **Requirement_6**) can be accomplished either directly via the Edge Routers 1103, 1104 (shown in Figure 11) and the BBA/BBD elements 1105, 1107, or, via the Edge Router ER1, the first Border Router BR1, and BBA/BBD elements 1105, 1107.

The management of the resources in the interior of each Diffserv domain, i.e., Core Routers is accomplished by each BR (see **Requirement_5**). This can be achieved by using a protocol such as Load Control (shown at 1111 in figure 11) described earlier.

Each BB (1105, 1106, 1107, communicates with all the ingress BR's (1112) of its Diffserv domain by using protocols 1118 e.g., Common Open Policy Service (COPS) or Simple Network Management Protocol (SNMP). In this way the BB is able to request from all the ingress BR's to either reserve a certain amount of resources or refresh a reservation that has been accomplished during a previous refreshment period.

Each ingress BR (1112) will use Load Control (1111) to reserve the resources requested by the BB. Afterwards, each BR will have to inform the BB about the amount of the resources that are actually reserved by the Load Control protocol. Each BB must contain a reservation state that will store the total amount of resources that were reserved by the Load Control protocol. This reservation state is only updated if either the BBA (1105) is requesting to modify it or if the resource conditions in the Diffserv network, i.e., Core routers, suddenly change.

As already explained earlier, the operation of the Load Control (1111) protocol is based on refreshment periods. If a certain amount of resources are reserved during a refreshment period, say period(i), then these resources will be used by the Diffserv domain during period(i+1). If these resources also have to be used during period (i+2), then these reservations have to be refreshed during period (i+1). It is proposed herein

that each BB in combination with the BRs will manage the refreshment procedure in its Diffserv domain. During each period the BB will use its reservation state information to find out how many units of resources, per RSVP aggregated session, will have to be refreshed during the next refreshment period.

In the proposed Intserv/Diffserv framework, three operation states are distinguished:

- Resource Reservation state: using among others the resource reservation state of the Load Control protocol the aggregated resources are reserved in all the BB's that are used in the end to end communication.
- Resource refreshment state: all the reserved resources in each Diffserv domain that have to remain reserved during the next refreshment period, have to be refreshed.
- Resource release state: all the reserved resources in each Diffserv domain that have to be released during the next refreshment period, will not be refreshed.

Resource reservation state

The operation of the RSVP aggregation protocol in the proposed Intserv/Diffserv framework during the Resource Reservation state depends among others on the availability of the RSVP aggregation states in the BB's. There are three situations that can be identified:

- Scenario_1: the BB's do not contain any RSVP aggregated state. Therefore, a scenario similar to the one explained in Figure 3 has to be used in order to create these RSVP aggregated states.
- Scenario_2: the BB's contain RSVP aggregated states and the new E2E RSVP request 1113 does not require the resizing of the RSVP aggregated states 1110. Therefore, a scenario similar to the one explained in Figure 4 has to be used.
- Scenario_3: the BB's contain RSVP aggregated states and the new E2E RSVP request 1113 requires the resizing of the RSVP aggregated states. Therefore, a scenario similar to the one explained in Figure 5 has to be used.

Scenario 1 (without aggregated states)

Figure 12 shows an exemplary Intserv/Diffserv operation, when RSVP aggregated states are not available in the Bandwidth Brokers (BBs). Figure 12 illustrates sender/edge router 1201, receiver/edge router 1202, ingress border routers 1212, eggress border routers 1214 and intermediate Diffserv domains 1209.

If the scenario 1 is applied to the network shown in figure 11, it is assumed that the RSVP aggregated states 1110 are not available in the BB's 1105, 1106, 1107 at the moment that a RSVP E2E message 1113 arrives at the BBA. The operation of the proposed Intserv/Diffserv for this scenario is based on the RSVP

aggregation operation viewed earlier in Figure 3, and it can be summarized as follows:

- Step_1: The Sender sends an RSVP E2E PATH message to the BBA. Note that this can be accomplished either directly via an Edge Router (shown in Figure 11) or via the Edge Router and the first Border Router that can communicate with a BBA and an Edge router.
- Step_2: The BBA 1205 receives the RSVP E2E PATH message 1213 and by using the IP tunneling encapsulation procedure or by using the new proposed in RSVP IGNORE option in [Balt00], it sends the RSVP E2E PATH message to the BBD 1207.
- Step_3: The BBD receives the RSVP E2E PATH message. The BBD depending on the number of the supported PHB's, sends one or more E2E PathErr messages to the BBA. In this example due to the fact that the BB's have to create two RSVP Aggregated states, one for an EF PHB and the other for an AF PHB, two such messages are sent.
- Step_4: The BBA after receiving these two E2E PATHErr messages, creates two RSVP aggregated states, one for each PHB. The BBA sends two RSVP AggPATH messages (A and B) to the neighboring BB. It is to be noted that in Figure 12, this BB is contained in the "Intermediate Diffserv domains" block. Note that the RSVP AggPATH messages have to be sent

through all the intermediate Diffserv domains, i.e., the BB's, that are providing the end to end communication.

- Step_5: The BBD receives these two AggPATH messages and sends the RSVP E2E message (after decapsulating it and adjusting its QoS specifications) to the Receiver 1202 (via an Edge Router).
- Step_6: The BBD by using SNMP or COPS will request from each ingress BR in its Diffserv domain to reserve the resources that were demanded by the RSVP AggPATH messages. Each BR will have to reserve a certain predefined percentage of the total amount of the resources that must be reserved.
- Step_7: Each ingress BR will use the Load Control mechanisms explained earlier to reserve the requested resources.
- Step_8: Each BR will send a Reply message to the BBD to inform the BBD about the status of this procedure. Note that the Reply message could be a SNMP or a COPS message.
- Step_9: The BBD using the obtained information from all the BR's, creates the aggregated reservation state and it sends two AggRESV messages to the neighboring BB. In Figure 12 this BB is contained in the "Intermediate Diffserv domains" block. Note that the RSVP AggRESV messages have to be sent

through all the intermediate Diffserv domains, i.e., the BB's, that are providing the end to end communication.

- Step_10: In each intermediate Diffserv domain (i.e., BB and BRs) and also in the initiating Diffserv domain (i.e., BBA and BRs) the functionality described in Step_6, Step_7, Step_8, and Step_9 has to be repeated. The difference is that in Step_6 the SNMP or COPS messages will request the resources that were demanded by the received RSVP AggRESV messages and not by the AggPATH messages.
- Step_11: The BBA sends via all the intermediate BB's two RSVP AggRESVConfirm messages to the BBD. The messages confirm the reservation of the resources.
- Step_12: The Receiver (via an Edge Router) replies with a RSVP E2E RESV message that among others contains the QoS parameters (specs) that can be supported by the Receiver.
- Step_13: The BBD encapsulates the RSVP E2E Resv message (using IP tunneling or the RSVP IGNORE option) that has been received from the Receiver (see Step_12) and sends it to the BBA.
- Step_14: The BBA decapsulates it and sends it to the Sender (via an Edge Router).

Scenario 2 (with aggregated states but without a need for resizing)

Figure 13 illustrates an exemplary proposed Intserv/Diffserv operation when RSVP aggregated states are

available in the BBs, and no resizing is needed. Shown in figure 13 are sender/edge router 1301, receiver/edge router 1302, intermediate diffserv domains 1309, ingress BRs 1312, aggress BRs 1314, BB Aggregator 1305 and BB deaggregator 1307.

In this Scenario as shown in Figure 13, it is considered that the RSVP aggregated states are available in the BB's at the moment that a RSVP E2E message arrives at the BBA. Furthermore, it is noted that the new E2E RSVP request does not require the resizing of the RSVP aggregated states. The operation of the proposed Intserv/Diffserv for this scenario is based on the RSVP aggregation operation illustrated in Figure 4 and it can be summarized as follows:

- Step_1: The Sender sends an RSVP E2E PATH message to the BBA. Note that this can be accomplished either directly via an Edge Router (shown in Figure 11) or via the Edge Router and the first Border Router that can communicate with a BBA and an Edge router.
- Step_2: The BBA receives the RSVP E2E PATH message and by using the IP tunneling encapsulation procedure, it sends the RSVP E2E PATH message to the BBD.
- Step_3: The BBD decapsulates the RSVP E2E PATH message and it sends it to the Receiver (via an Edge Router).

- Step_4: The Receiver replies (via an Edge Router) with a RSVP E2E RESV message that among others contains the QoS parameters that can be supported by the Receiver.
- Step_5: The BBD encapsulates the RSVP E2E Resv message (using IP tunneling or the RSVP IGNORE option) that has been received from the Receiver and sends it to the BBA.
- Step_6: The BBA decapsulates it and sends it to the Sender (via an Edge Router).

<u>Scenario 3 (with aggregated states but with a need for resizing</u>

In the Scenario shown in Figure 14, it is considered that the RSVP aggregated states are already available in the BB's and that the E2E RSVP request requires the resizing of these states. The operation of the proposed Intserv/Diffserv framework for this scenario is based on the operation viewed in Figure 5 and can be summarized as follows:

- Step_1: The Sender 1401 sends an RSVP E2E PATH message to the BBA 1405. Note that this can be accomplished either directly via an Edge Router (shown in Figure 11) or via the Edge Router and the first Border Router that can communicate with a BBA and an Edge router.

Figure 14 illustrates an example of a proposed Intserv/Diffserv operation when RSVP aggregated states are available in the BBs, and resizing is needed. Shown in figure

14 are sender/edge router 1501, receiver/edge router 1402, ingress BRs 1412, eggress BRs 1414, intermediate Diffserv domains 1409, BB Aggregator 1405 and BB Deaggregator 1407.

- Step_2: The BBA receives the RSVP E2E PATH message and by using the IP tunneling encapsulation procedure or by using the new proposed RSVP E2E IGNORE option in [Balt00], it sends the RSVP E2E PATH message to the BBD 1407.
- Step_3: The BBD decapsulates the RSVP E2E PATH message and it sends it to the Receiver 1402 (via an Edge Router).
- Step_4: The Receiver 1402 (via an Edge Router) replies with a RSVP E2E RESV message that among others contains the QoS parameters that can be supported by the Receiver.
- Step_5: The BBD 1407, in this example, will find out that the RSVP aggregated states are not enough to support the requested QoS parameters that are contained in the RSVP E2E RESV message. Therefore, the BBD 1407 will initiate a RSVP aggregated reservation resizing procedure.
- Step_6: The BBD 1407 by using SNMP or COPS will request from each ingress BR in its Diffserv domain to reserve the resources that were demanded by the RSVP E2E RESV message. Each BR will have to reserve a certain predefined percentage of the total amount of the resources that must be resized.

- Step_7: Each ingress BR will use the Load Control mechanism (described earlier), to resize the requested resources.
- Step_8: Each BR will send a Reply message to the BBD to inform it about the status of this procedure. Note that the Reply message could be a SNMP or a COPS message.
- Step_9: The BBD using the obtained information from all the BR's, resizes the aggregated reservation state and it sends one AggRESV message to the neighboring BB. In Figure 14 this BB is contained in the "Intermediate Diffserv domains" block. Note that the RSVP AggRESV message has to be sent through all the intermediate Diffserv domains, i.e., the BB's, that are providing the end to end communication.
- Step_10: In each of the intermediate Diffserv domains (i.e., BB and BRs) and in the initiating Diffserv domain (i.e., BBA and BRs) the functionality described in Step_6, Step_7, Step_8 and Step_9 has to be repeated.
- Step_11: The BBA sends via all the intermediate BB's one RSVP AggRESVConfirm message to the BBD. This message confirms the resizing of the reserved resources.
- Step_12: The BBD encapsulates the RSVP E2E Resv message (using IP tunneling or the RSVP IGNORE option) that has been received from the Receiver (see Step_4) and sends it to the BBA.

- Step_13: The BBA decapsulates it and sends it to the Sender (via an Edge Router).

Resource refreshment state

Due to the fact that the proposed Intserv/Diffserv framework is using the Load Control protocol to reserve the Aggregated resources within each Diffserv domain, it has also to refresh these resources during each refreshment period.

It is noted that the refreshment procedure in each Diffserv domain should be managed by the BB that is managing the QoS in the Diffserv domain in combination with its ingress BR's. In each refreshment period and for each RSVP aggregation state the BB in each Diffserv domain will have to inform each BR about the number of the resources that have to be refreshed. Using the Load Control protocol each BR will then refresh the reservation of the resources reserved. This operation is described earlier and it can be achieved by sending for each unit of resource one RP packet. This operation is generally illustrated in Figure 15, which is an example of refreshment of the reserved resources, and shows sender/edge router 1501, receiver/edge router 1502, ingress BRs 1512, eggress BRs 1514, intermediate Diffserv domains 1509, BB Aggregator 1505 and BB Deaggregator 1507.

Resource release state

As explained earlier, the E2E RSVP reservation states are temporary states, i.e., soft states, that have to be updated temporarily. This means that E2E PATH and E2E RESV messages will have to be periodically retransmitted. If the states are not refreshed then they will be removed. These states may also be removed by using the E2E PATHTear and E2E RESVTear messages. The refreshment, update and release of the aggregated states is based on a certain predefined policy which the Aggregator and Deaggregator will decide when the RSVP Aggregated states will be refreshed or updated; this triggering time is not completely defined by the E2E RSVP messages. In particular, this predefined policy takes into account the sum of the underlying E2E reservations, and a certain level of trend analysis.

Within each Diffserv domain the release of the resources is managed by each BR and is accomplished by using the Load Control protocol. If the BR does not receive any request for change in its reserved resources from the BB, then it will assume that it will have to release all the resources that it is managing. The BR will release a previously reserved resource by not refreshing it.

ADVANTAGES OF THE INVENTION

This invention offers a novel concept and method that can be used to combine an Intserv region(s) with a dynamically

provisioned Diffserv domain, where the QoS management is controlled by a new type of Bandwidth Brokers. This approach enhances and extends the Intserv over Diffserv framework, i.e., Solution_4 described supra. The advantages of this new concept compared to the one described as Solution_4 given earlier are:

- The BB is able to directly communicate and manage only the Border Routers and not the Core Routers in the Diffserv Domain. The Border Routers will manage the resource availability and the admission control into the interior of the Diffserv domain, i.e., on the Core Routers. This can be achieved by using the Load Control protocol specified earlier. In this way the dynamic QoS provisioning in Diffserv architectures does not impose severe scalability problems on the BB and on the Core Routers of the Diffserv domain. Furthermore, due to the fact that the RSVP aggregated states are now only stored into the BB, the problem related to large full meshed networks will not anymore occur. In other words **Issue_2** described earlier can be efficiently solved.
- The interoperation between the BB's of each Diffserv domain can be accomplished quite efficiently and easily. Therefore, **Issue_3** described above is efficiently solved.

It is important to note that the combination of the RSVP aggregation and the Load Control concepts can be also used when

RSVP is not applied end to end. In this case the Aggregator can use a policy that can be based on local configurations and local QoS management architectures, to set the DSCP packets that are passing into the aggregated region. This means that this concept can also be applied on e.g., PSTN (Public Switched Telephone Networks), GPRS (General Packet Radio Service) or UMTS (Universal Mobile Telecommunications System) networks that are using the Diffserv concepts in their Core networks.

EQUIVALENTS

Although preferred embodiments of the method and apparatus of the present invention have been illustrated in the accompanying drawings and described in the foregoing detailed description, it will be understood that the invention is not limited to the embodiments disclosed, but includes numerous rearrangements, modifications, equivalents and substitutions without departing from the scope of the invention as set forth in the appended claims.

List of References:

[BeBi99] Bernet, Y., Binder, J., Blake, S., Carlson, M., Davies, E., Ohlman, B., Werma, D., Wang, Zh., Weiss, W., "A Framework for Differentiated Services", draft-ietf-diffserv-framework-02.txt, February 1999.

[RFC2998] Bernet, Y., Yavatkar, R., Ford, P., Baker, F.,Zhang, L., Speer, M., Braden, R., Davie,B., felstaine, E. "Framework for Integrated Services operation over Diffserv Networks",draft-ietf-issll-diffserv-rsvp-04.txt, March 2000 (work in progress).

[DuYa99] Durham, D., Yavatkar, R., "Inside the Internet's Resource reSerVation Protocol: Foundations for Quality of Service", John Willey & Sons, Inc., 1999.

[RFC1633] Braden, R., Clark, D., Shenker, S., "Integrated Services in the Internet Architecture: An Overview", IETF RFC1633, 1994.

[RFC1905] Case, J., McCloghrie, K., Rose, M., and S. Waldbusser, "Protocol Operations for Version 2 of the Simple Network Managment Protocol (SYMPv2)", RFC 1905, 1996.

[RFCC2205] Braden, R., Zhang, L., Berson, S., Herzog, S., Jamin, S., "Resource Reservation Protocol (RSVP) Version 1 Functional Specification", IETF RFC 2205, 1997.

[RFC2210] Wroclawski, J., "The use of RSVP with IETF integrated Services", IETF RFC 2210, 1997.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Zh., Weiss, W., "An architecture for Differentiated Services", IETF RFC 2475, 1998.

[RFC2597] Heinanen, J., Baker, F., Weiss, W., Wroclawski, J., "Assured Forwarding PHB group", IETF RFC 2597, 1999.

[RFC2598] Jacobson, V., Nichols, K., Poduri, K., "An Expedited Forwarding PHB", IETF RFC 2598, 1999.

[RFC2638] Nichols, K., Jacobson, V., Zang, L., "A two-bit Differentiated Services Architecture for the Internet", IETF RFC 2638, 1999.

[RFC2748] Durham, D., Boyle, J., Cohen, R., Herzog, S., Raja, R., Sastry, A., "The COPS (Common Open Policy Service) Protocol", IETF RFC, January 2000.